The Text/HTML content type and the Content-Location MIME header

or

Sending HTML documents via MIME e-mail

Status of this Memo

This document is an Internet-Draft. Internet-Drafts are working
documents of the Internet Engineering Task Force (IETF), its
areas, and its working groups. Note that other groups may also
distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six
months and may be updated, replaced, or obsoleted by other
documents at any time. It is inappropriate to use Internet-
Drafts as reference material or to cite them other than as
"work in progress."

To learn the current status of any Internet-Draft, please check
the "$id-abstracts.txt" listing contained in the Internet-
Drafts Shadow Directories on ftp.is.co.za (Africa),
nic.nordu.net (Europe), munnari.oz.au (Pacific Rim),
ds.internic.net (US East Coast), or ftp.isi.edu (US West Coast).

This memo provides information for the Internet community. This'
memor does not specify an Internet standard of any kind, since
this document is mainly a compilation of information taken from
other RFC-s. Distribution of this memo is unlimited.

Abstract

This memo specifies how to send HTML-formatted documents in Internet
mail. The memo particularly addresses the issue of handling of
hyperlinks in HTML documents referring to other body parts in the same
message. In order to do this, the memo introduces one new MIME content-
header with the name "Content-Location" and one new attribute to the
MIME header "Content-Type: Text/HTML" with the name "linking".
1. Introduction

The HTML format is a very common format for documents in the Internet, and there is an obvious need to be able to send documents in this format in e-mail [SMTP, RFC822]. The "text/html; version=2.0" media type is defined in [HTML2]. This memo gives additional specifications and advice on how to use the text/html media type as a Content-Type in MIME [MIME] e-mail messages.

2. Terminology

Most of the terms used in this memo are defined in other RFC-s. For example, URL is defined in [URL], URI, absolute URI, and relative URI is defined in [HTML2].

3. The Content-Location MIME content-header

An additional MIME heading field is defined with the name "Content-Location". This header field can occur in any MIME message heading or content heading. Its value is an absolute URI. The data below the current heading, and the data retrievable through this URI, should be identical. In practice, at present only those URI-s which are URL-s are used, but it is anticipated that other forms of URI-s will in the future be used. This heading is similar to the Location header as defined in [HTTP].
The syntax for the new heading field is, using the syntax definition tools from [RFC822]:

\[
\text{content-location ::= "Content-Location:" <> uri <>}
\]

where uri is at present (November 1995) restricted to the syntax for URL-s as defined in [URL]. This syntax will be widened when the definition of URI syntax becomes more stable.

Question: Is it necessary with the quotes (<>>) around the value? HTTP does not have such quotes around the value of the Location heading field in HTTP.

4. Parameters for the Content-Type: Text/HTML

The optional "version" parameter for the Content-Type: Text/HTML indicates the version of HTML used, with "2.0" as default value.

The new optional parameter "linking" defined in chapter 10 below.

5. Use of relative URL-s in Text/HTML contents

The use of relative URL-s in Content-Type: Text/HTML SHOULD never be used except in one of the following three cases:

(a) There is a BASE element in the HTML document which resolves the relative URL into a non-relative URL.

(b) The relative URL refers to another body part in the same message as defined in chapter 9.2 Filename-method below.

(c) In special cases, where the sender and recipient have prior agreement on the resolution of relative URL-s.

6. Use of the Content-Type: multipart/related

A message can contain one or more Text/HTML body parts and also contain as separate body parts, data, to which hyperlinks (as defined in [HTML2]) in the Text/HTML body part refers. If this is done, it is recommended to use the multipart/related Content-Type as defined in [REL]. The root (as defined in [REL]) should then be of the Content-Type: Text/HTML.

7. Use of Content-type: Multipart/alternative

If the message is sent to recipients, all of which may not have mailers capable of handling the Text/HTML content-type, then the Content-Type: Multipart/Alternative [MIME] can be used, for example with Content-Type: Text/plain as the first choice, and Content-Type: Text/HTML as the second choice.
8. Combination of Content-Type: Multipart/related and Multipart/alternative.

Both the Content-type: Multipart/related, as defined in chapter 6 above and the Content-Type: Multipart/alternative, as defined in chapter 7 above can be combined in the same message. It is then recommended to put the Multipart/alternative inside the Multipart/related.

Example:

```
Content-Type: Multipart/related; boundary="boundary-example-1"
  type=Text/HTML

--boundary-example-1
Content-Type: MULTIPART/ALTERNATIVE
Boundary: boundary-example-2

--boundary-example-2
Content-Type: Text/plain

... plain text version of the document for recipients whose mailers cannot handle Text/HTML ...

--boundary-example-2
Content-Type: Text/HTML

... text of the HTML document ...

--boundary-example-2--
--boundary-example-1
Content-Type: Image/GIF

... a body part, to which the HTML document has a link ...
--boundary-example-1--
```

9. Links to other body parts

A Text/HTML body part may contain hyperlinks to documents which are included as other body parts in the same message. Three ways to do this is specified in this memo:

9.1 Location-method: Use of the Content-Location field

With this method, All URI-s in the Text/HTML document SHOULD be non-relative URI-s as defined in [HTML2], and it SHOULD be possible to use these URI-s to retrieve the referred document using the protocol defined for retrieval of this particular URL scheme in [URL] (subject to access control).
For each distinct URI in the Text/HTML document, which refers to data which is sent in the same MIME message, there SHOULD be a separate body part in the message containing this data. Each such body part SHOULD contain a Content-Location heading field, and the value of this field SHOULD be identical to the URI as used in the Text/HTML document.

The receiving mailer can then resolve the hyperlink either by using the URI in the normal way, or by using the data in the body part whose Content-Location contains the same URI.

Example:

Content-Type: Multipart/related; boundary="boundary-example-1"; type=Text/HTML

--boundary-example 1
Content-Type: Text/HTML

... text of the HTML document, which might contain a hyperlink to the other body part, for example through a statement such as:
<IMG SRC="http://www.dsv.su.se/images/DSV-sign-eng.gif">

--boundary-example-1
Content-Type: Image/GIF
Content-Location: "http://www.dsv.su.se/images/DSV-sign-eng.gif"

--boundary-example-1--

9.2 Filename-method: Use of file names

With this method, the hyperlink URIs to other body parts in the same message in the Text/HTML document SHOULD have a very simple format. This simple format is relative URL-s of the form

relative-url ::= 1ALPHA 0#7ALPHADIGIT [ "." 1#3ALPHADIGIT ]

ALPHADIGIT ::= ALPHA / DIGIT

i.e. 1-8 characters plus 0-3 extension characters, only using Ascii letters and digits and beginning with a letter.

The choice of this simple format is to match permitted file name formats in most operating systems in wide use today.

For each distinct URI in the Text/HTML document, which refers to data which is sent in the same MIME message, there should be a separate body part in the message containing this data. Each such body part SHOULD contain a Content-Disposition header [RFC 1806] with a filename parameter. The value of this filename parameter should be identical to the relative URI as used in the Text/HTML document.
The value of the Content-Disposition header should be "inline" if the URI in the Text/HTML document is used for an inline HTML element, such as an <IMG> element, and should be "attachment" if the URI in the Text/HTML document is used for a hyperlink to a document which is activated at the request of the recipient, such as an <A> element.

Example:

```
Content-Type: Multipart/related; boundary="boundary-example-1"; type=Text/HTML

--boundary-example 1
Content-Type: Text/HTML

... text of the HTML document, which might contain a hyperlink to the other body part, for example through a statement such as:
<IMG SRC="signeng.gif">

--boundary-example-1
Content-Type: Image/GIF
Content-Disposition: inline/ filename=signeng.gif

--boundary-example-1--
```

9.3 CID-method: Use of CID URL-s

With this method, the hyperlink URIs to other body parts in the same message in the Text/HTML document SHOULD be CID (Content-ID) URL-s as defined in [URL] and [MIDCID].

For each distinct URI in the Text/HTML document, which refers to data which is sent in the same MIME message, there should be a separate body part in the message containing this data. Each such body part SHOULD have a Content-ID header [MIME]. The value of this Content-ID header should be identical to the CID as used in the Text/HTML document.

Example:

```
Content-Type: Multipart/related; boundary="boundary-example-1"; type=Text/HTML

--boundary-example 1
Content-Type: Text/HTML

... text of the HTML document, which might contain a hyperlink to the other body part, for example through a statement such as:
<IMG SRC="cid:sign-eng*jpalme@dsv.su.se">

--boundary-example-1
Content-Type: Image/GIF
Content-ID: sign-eng*jpalme@dsv.su.se

--boundary-example-1--
```
9.4 Recommended choice of method:

A Text/HTML content may always, in addition to the use the methods described in this chapter of this memo, contain URI-s only resolvable using the method defined for this particular URI scheme, and not referring to any data in separate body parts of the same message.

<table>
<thead>
<tr>
<th>Method</th>
<th>Body part identification method</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location-method</td>
<td>Content-Location</td>
<td>When the referred document is publicly available and retrievable using the scheme used in the URI.</td>
</tr>
<tr>
<td>Filename-method</td>
<td>filename in Content-Disposition header</td>
<td>For private documents or documents not retrievable.</td>
</tr>
<tr>
<td>CID-method</td>
<td>Content-ID</td>
<td>For experimental use between consenting partners.</td>
</tr>
</tbody>
</table>

10. Indication of method used

There should be an additional optional parameter to the Content-Type: Text/HTML header, with the name "linking" and the syntax:

linking ::= "linking=" linkmethod

where linkmethod can have the following values:

- **external**: Only use of absolute URI-s and retrieval using the scheme defined for this URI, and not containing any URI referring to other body parts in the same message.

- **location**: Location-method as defined in chapter 9.1.

- **filename**: Filename-method as defined in chapter 9.2.

- **cid**: CID-method as defined in chapter 9.3.

Default value if this parameter is omitted is "external".
11. Encoding considerations

There are two recommended ways to encode 8-bit characters in Text/HTML contents:

(1) Let the charset of the content part be iso-8859-1, and encode the content with the quoted-printable encoding method.

(2) Let the charset of the content part be us-ascii, and encode non-us-ascii characters in the text using the Data character encoding defined in [HTML2].

Both these encoding methods are permitted, and they can also be mixed in the same document. Recipients must be capable of handling both encoding alternatives. However, it is recommended that encoding method (2) above is used when sending Text/HTML messages.

If only method (2) is used, the charset parameter should be "us-ascii".

If method (1), or a mixture of method (1) and method (2) is used, the charset parameter should be "iso-8859-1".

12. Security considerations

There is a potential security risk if the Content-Location: heads a body part whose data is not identical to that retrievable using the URI in the Content-Location. To reduce this risk, it might be unsuitable to cache the data in such a way that the cached data can be used for retrieval of this URL from other documents than those included in the same message as the Content-Location header.

13. Acknowledgments

Harald Tveit Alvestrand, Keith Moore, Ed Levinson, Al Gilman, Valdis Kletnieks, Larry Masinter and several other people have helped me with preparing this memo. I alone take responsibility for any errors which may still be in the memo.

14. References

Temporary note: This list contains some references to Internet drafts. It is anticipated that these Internet drafts will become RFC-s before this memo. The references will then in this memo be changed to refer to the corresponding RFC instead.

Ref.         Author, title
------------ --------------------------------------------------------


15. Author’s address

Jacob Palme                          Phone: +46-8-16 16 67
Stockholm University and KTH         Fax: +46-8-783 08 29
Electrum 230                         E-mail: jpalme@dsv.su.se
S-164 40 Kista, Sweden